Mathematics publications and authors' gender: Learning from the Gender Gap in Science project

Helena Mihaljević and Lucía Santamaría

The number of women pursuing higher education in STEM areas (science, technology, engineering and mathematics) is steadily growing, yet their presence progressively decreases when it comes to high-level academic positions. The COVID-19 crisis has aggravated preexisting gender inequalities so that, according to the most recent Global Gender Gap Report of the World Economic Forum, "another generation of women will have to wait for gender parity" [9]. The pandemic has hit female academics in STEM particularly hard along multiple dimensions, such as productivity, boundary setting and control, as well as the ability to engage in collaborations and network building [6]. In order to fully understand the gender gap in academia and its development, for instance to assess and counteract the effects of crises such as pandemics, fine-grained data are needed, but these are unfortunately not generally collected and analyzed in a consistent manner.

1 The Gender Gap in Science project

The three-year project "A Global Approach to the Gender Gap in Mathematical, Computing, and Natural Sciences: How to Measure It, How to Reduce It?"[1] was funded by the International Science Council (ISC) in 2017–2020. It brought together eleven scientific organizations, led by the International Mathematical Union (IMU) and the International Union of Pure and Applied Chemistry (IUPAC). The goal of this interdisciplinary, cross-national project was to study the situation of women in mathematics, computing and natural sciences. The work was articulated around three central themes aiming to identify common and discipline-specific issues that might require interventions. Two of the tasks were the launch of a Global Survey of Scientists and the creation of a Database of Good Practices. Moreover, a data-driven study was conducted in order to *examine the situation of academic authors and their publication practices in different fields across world countries and regions with respect to the scientists' gender.*

The interest in bibliometric analyses is rooted in the importance of published papers for academic careers. Scientific publications are not only the major outlet for scholarly communication, they are regarded as a proxy for a researcher's scientific credibility and play a key role in achieving and maintaining a successful career in academia. Decisions on tenure and other academic promotions are mostly based on evaluations of the candidate's research portfolio that pay special attention to publications such as journal articles, in addition to grants, conference presentations, and how visible or well-recognized a scholar is. Thus, *the understanding of publication practices, obtained through measurable data on research output, is of great interest* to academic institutions, science policymakers, and researchers alike.

Multiple studies based on bibliometric data have focused on gender. The literature even comprises discipline-specific findings pertaining to mathematics and physics, albeit in small numbers. Much of the existing scientometric research builds on cross-discipline corpora such as Scopus and accordingly does not highlight individual fields. Research with a topical focus on a particular discipline or subfield is typically limited to a selection of journals or conferences or a narrow time period. For the work executed within the Gender Gap project, we decided to analyze the most comprehensive data sources in terms of content and temporal coverage. Those collections happen to be managed by community organizations and curated by experts: data for mathematics came from zbMATH; for theoretical physics, we used arXiv preprints enriched with Crossref; for astronomy and astrophysics, we resorted to ADS.

For the field of mathematics, we analyzed zbMATH's full collection of publications by authors with a main research focus in mathematics from 1970 until July 2019. This *data set comprises more than 3 million documents corresponding to more than 5.2 million authorships* (pairs of author and document), yielding an average of 1.7 authors per article. We inferred the gender of these authorships from the authors' names via various statistical name-gender databases and services. The resulting gender breakdown was approximately 70 % men, 10 % women, and 20 % undetermined. Omitting the latter, women accounted for about 12 % of the male plus female authorships. These[2] in turn belong to ca.

---

[2] Not all authorships can be assigned to a unique author, in particular if the author's name is frequent.

65,000 and ca. 260,000 distinct authors labeled as women and men, respectively, which yields around 21 % distinct women among all recorded authors in zbMATH in the past 50 years, growing from less than 10 % in the 1970s to over 27 % today. Currently, the rate of *new authors being added to the database is more than 14,000 per year, which means that ca. 4,000 women enter the field of mathematics annually.*

One key aspect of publishing relates to journals and their perceived quality. In mathematics, research is predominantly driven by scholarly journals, and they are a crucial vehicle for the forging of academic careers. Publishing in highly renowned venues is a powerful determinant of tenure and an important predictor of professional success. Therefore, any bibliometric study on publication practices has to take into account their impact in the making of academic careers. Here we present some mathematics-related findings from the work done within the Gender Gap in Science project. Further results plus additional context information, e.g. on the data processing algorithms that were employed, can be found in the final project report [7].

Representation of women in renowned mathematical journals

Previous research [5] showed that authorships by women in mathematics are vastly underrepresented in journals with a high reputation in terms of two common ranking methods, the manually compiled Australian ERA indicator and the Thomson Reuters journal impact factor (JIF). In this project, we intended to offer the scientific community the opportunity of examining gender distributions in journals of particular relevance to them or their subfield. We made this possible via a dedicated web interface³ that allows readers to filter specific publication venues of their interest.

In addition, we have specifically analyzed *selected journals of particular renown in concrete subfields as well as several journals published by mathematical societies*. Figure 1 illustrates the percentages of authorships from women in said selected journals, which are predominantly constrained below 20 %. Only half of the society journals show a rising tendency over the past decades. No evident increase in women representation can be seen in the *Bulletin de la Société Mathématique de France* or the *Journal of the European Mathematical Society*, with both stagnating below 10 %. Even lower, at 5 % and with no sign of improvement, is the presence of women in the *Journal of the American Mathematical Society*. When it comes to discipline-specific topics, as is the case of the three journals on the bottom right that focus on applied mathematics, we observe a rising development and at least 10 % of women's presence. Apart from the *Journal of Differential Geometry*, all specialized journals reveal a moderate positive trend. The prestigious journals in pure mathematics *Inventiones Mathemat-*

icae and *Annals of Mathematics* stand out with percentages of women authorships predominantly in the single-digit range. For more details, see [4].

Several factors can be hypothesized to be potential causes for the measured underrepresentation, but these cannot be confirmed by the bibliographic data alone. As a complementary data source, we have leveraged the 2018 Global Survey of Mathematical, Natural, and Computing Scientists that was conducted within the Gender Gap Project. The survey reached almost 10,000 mathematicians, physicists, and astronomers who were questioned about their submission practices to top-ranked journals in their disciplines. More precisely, the following was asked: "During the last five years, how many articles have you submitted to journals that are top-ranked in your field?" Respondents were expected to provide a number between 0 and 30; larger values were clustered together. According to the responses obtained, women and men self-report to have submitted similar numbers of articles in the past 5 years, with no statistically significant differences in subgroup analyses broken down by disciplines or world regions.

In order to determine the most important predictors for the number of paper submissions, we trained a statistical model that took into account not only the gender but also other factors like discipline, country, access to childcare, as well as various aspects of the academic career and the professional activity, such as the number of grant applications and supervised graduate students. We found out that gender plays a minor role in the model; far more important are aspects that signal career advancement such as having a broad network and strong research activity, which are indirectly linked to gender to some extent.

In conclusion, the self-report of perceived submission practices does not support the hypothesis that women are underrepresented in prestigious journals because they submit fewer manuscripts than men. Considering the importance of publications in renowned journals on the one hand and the conflicting bibliometric analysis on the other, this prompts the question of the role of the peer review process. We note that the refereeing system in mathematics lacks homogeneity and relies substantially on the authors' credit and the level of trust between editors and reviewer(s). In this regard, we stress that *there are hardly any systematic studies on peer review in mathematics* [1], a need that very much ought to be addressed.

By publishing (analyses of) acceptance rates broken down by gender and other sociodemographic or career-related aspects, publishers of high-impact journals could make an important contribution to the evaluation of the fairness of the publication process. Self-assessed data by one of the major publishers in physics suggests at least a gender- and workplace-based bias in physics [3]. Unfortunately, this type of study is far from being a common practice among scientific publishers. A noteworthy exception is *The British Journal for the Philosophy of Science*, which provides acceptance rates broken down by author gender and affiliation

---

³ https://gender-publication-gap.f4.htw-berlin.de/

Figure 1. Percentage of authorships from women in renowned mathematics journals per year between 1970 and 2017 [4].
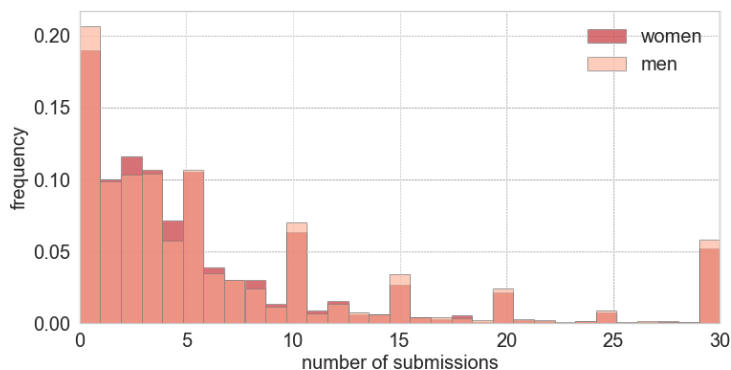
Figure 2. Histogram of the number of publications submitted to top-ranked journals in the last 5 years as self-reported in the Global Survey. Dark and light bars encode answers from women and men, respectively [5].

country. Interestingly, papers by women as sole authors have higher acceptance rates [8].

In relation to the impact of COVID-19 in publication practices, the journal Isis from the History of Science Society has evaluated its submission and acceptance rates, noting an "alarming drop in manuscript submissions by female scholars" in the first half of 2020 [2]. A literature search for similar evaluations of mathematical publishers remains unsuccessful. Along the same lines, the search for studies on the effects of the pandemic for female mathematicians only leads to results based on preprint repositories such as the arXiv, indicating the absence of data from publishers' self-assessments.

Learning and perspectives

There are various aspects to consider when speaking of a gender gap. In the ISC Gender Gap project, we have provided insights on the gap defined by the proportional presence of women as authors of core mathematics publications, we have investigated whether there is a gender gap in the dropout rates that affect the span of mathematicians' academic careers, and we have focused on the gender gap in high-impact mathematical journals.

Consistent with the global trend in higher education, we observe *increasing proportions of women entering the field of mathematics with each passing year*. The understanding of the extent to which those newcomers will progressively attain senior academic positions is crucial to address the "leaky pipeline" phenomenon. Thanks to our cohort analysis based on zbMATH publication data, we are able to provide insights on this issue. We show that dropout rates of mathematicians after their postdoctoral stage, which used to be higher for women than for men, are progressively converging. These data certainly offer optimistic prospects regarding the eventual closure of this particular aspect of the gender gap.

On the other hand, our analysis of women's presence in renowned journals is a good measure of the gender gap in relation to the achievement of a prestigious academic career. In this regard, a non-negligible number of *the prominent mathematical journals under consideration show a meager representation of female authors*. All other factors being equal, the expectation is that the proportion of women among all authors should roughly resemble the percentage of established female mathematicians in the profession, a number that has been steadily growing and that is estimated to be around 25 %. Remarkably, several of the analyzed journals publish very few articles authored by women and have exhibited no signs of turnaround over the last couple of decades. An explanation for this fact might lie in the characteristics of the peer review process in mathematics, which favors close interactions and trust relationships between editors and reviewers and opens the door to conscious and unconscious biases. Regarding subfields, applied areas display a better situation for women than pure ones, which in itself introduces a number of questions regarding the intrinsic differences among subfields of mathematics.

The above remarks provide a compelling starting point for future research. Is the increasing number of young female mathematicians enough to stop the pipeline from leaking? Which factor in the retention of women in academia is played by the professional atmosphere in pure versus applied mathematics? What is the importance of informal academic networks in helping a mathematician's career to thrive? Is the lack of double-blindness in peer review hindering women and other underrepresented groups in mathematics? *It would be excellent to discuss our data-backed findings with experts from the respective subfields in the mathematical community*, with the goal of formulating plausible hypotheses that could explain the observations found by our work in the Gender Gap project.

References

[1] C. Geist, B. Löwe and B. V. Kerkhove, Peer review and knowledge by testimony in mathematics. In *PhiMSAMP. Philosophy of Mathematics: Sociological Aspects and Mathematical Practice. Including Selected Papers of the 3rd PhiMSAMP Conference "Is mathematics special?" (Vienna, 2008)*, edited by B. Löwe et al., Texts in Philosophy 11, College Publications, London, 155–178 (2010)

[2] History of Science Society, Updated report on Isis submissions and gender during the pandemic, hssonline.org/report-on-isis-submissions-and-gender-during-the-pandemic/

[3] IOP Publishing, Diversity and inclusion in peer review at IOP Publishing, IOP Publishing, ioppublishing.org/wp-content/uploads/2018/09/J-VAR-BK-0818-PRW-report-final2.pdf (2018)

[4] H. Mihaljević and L. Santamaría, Authorship in top-ranked mathematical and physical journals: Role of gender on self-perceptions and bibliographic evidence, *Quant. Sci. Stud.* **1**, 1468–1492 (2020)

[5] H. Mihaljević-Brandt, L. Santamaría and M. Tullney, The effect of gender in the publication patterns in mathematics. *PLOS ONE* **11**, e0165367 (2016)

[6] National Academies of Sciences, Engineering and Medicine, *The impact of COVID-19 on the careers of women in academic sciences, engineering, and medicine*. The National Academies Press, Washington, DC (2021)

[7] M.-F. Roy et al., *A global approach to the gender gap in mathematical, computing, and natural sciences: how to measure it, how to reduce it?* DOI 10.5281/zenodo.3882609 (2020)

[8] The British Journal for the Philosophy of Science, BJPS Stats 2020. BSPS, thebsps.org/auxhyp/bjpsstats2020/ (2021)

[9] World Economic Forum, *Global gender gap report 2021*. World Economic Forum, www3.weforum.org/docs/WEF_GGGR_2021.pdf (2021)

Helena Mihaljević is professor of data science at the Hochschule für Technik und Wirtschaft (HTW) Berlin, in association with the Einstein Center Digital Future in Berlin. The focus of her teaching and research activities is on statistical data analysis, applied machine learning, and information retrieval. Previously, she worked as a senior data scientist at The Unbelievable Machine Company in Berlin, as editor and deputy head at zbMATH, and as a research associate at Christian-Albrechts-Universität zu Kiel. She studied mathematics at the Georg-August-Universität Göttingen and received her PhD in mathematics from the University of Liverpool in 2009 on topological dynamics of entire transcendental functions.

helena.mihaljevic@htw-berlin.de

Lucía Santamaría is an applied scientist and instructor at Amazon's Machine Learning University, where she develops and teaches a broad range of ML courses, from mathematical foundations to deep learning methods. One of her research interests is the study of the gender gap in STEM, following her prior work in named entity recognition and author identification at zbMATH. Lucía holds a PhD in physics from the University of Potsdam for her research on the modelling of black hole binaries as sources of gravitational-wave searches with LIGO, conducted at the Max Planck Institute for Gravitational Physics.

lucia.santamaria@ymail.com